

# THE GENOME OF *ARABIDOPSIS*

In December of 2000, the *Arabidopsis* research community announced a major accomplishment: the completion of the sequence of a flowering plant. For the first time, we have in hand the sequence of all of the genes necessary for a plant to function, knowledge unprecedented in the history of science. Additionally, this sequence is freely available to every member of the scientific community. Below is a summary of major findings described in a groundbreaking paper, "The *Arabidopsis* Genome Initiative", *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana*" (*Nature* 408, 796-815, 2000).

The genome of *Arabidopsis*:

- Contains about 125 megabases of sequence
- Encodes approximately 25,500 genes
- Contains a similar number of gene functional classifications as other sequenced eukaryotic genomes (*Drosophila melanogaster* and *Ceanorhabditis elegans*) (see Figure 1)
- Has 35% unique genes
- Has 37.5% genes that exist as members of large gene families (families of 5 or more members)
- Shows evidence of ancient polyploidy: an estimated 58-60% of the *Arabidopsis* genome exists as large segmental duplications (see Figure 2)

Figure 1: Functional analysis of *Arabidopsis* genes – Proportion of predicted *Arabidopsis* genes in different functional categories. Copyright Nature Publications (<http://www.nature.com/>), as published in *Nature* 408, 796-815 (14 December 2000). Permission to use kindly granted by Nature Publications.

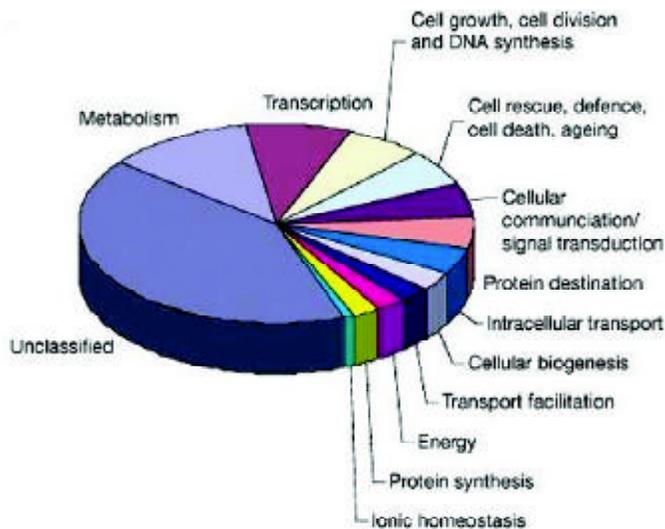
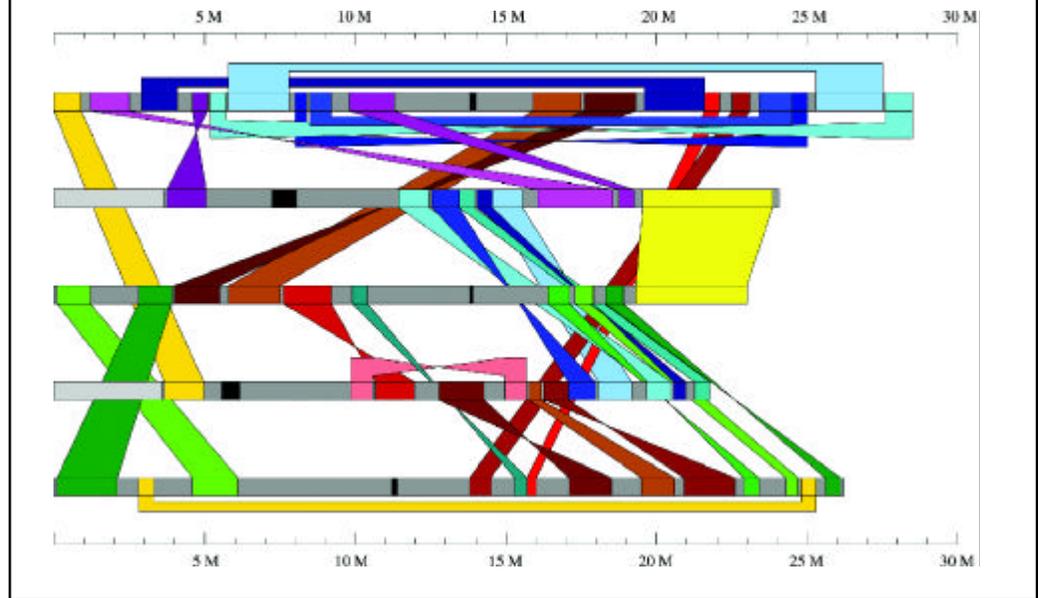


Figure 2: Summary of segmental duplications in *Arabidopsis thaliana*. The five chromosomes are depicted as grey horizontal bars, centromeres marked black. Colored bands serve as representations of duplicated segments. Bands are twisted if the corresponding segments have reversed orientation. The scale is in Megabasepairs (Mbp). Courtesy of Dirk Haase of MIPS (<http://mips.gsf.de>), and copyright Nature Publications (<http://www.nature.com/>), as published in Nature 408, 796-815 (14 December 2000). Permission to use kindly granted by Nature Publications.



Analysis of the sequence of the *Arabidopsis* genome tells us that the genome of a higher plant is similar in several important ways to the genome of other sequenced multicellular organisms. It also points out several important differences, which may not be too surprising, considering that plants differ in many important ways from the animals whose genomes have been analyzed. Plants are autotrophic: they require only light, water, air and minerals to survive. They can therefore be expected to have genes that animals do not have, encoding the proteins and enzymes involved in plant-specific processes, including the complex process of photosynthesis.

- *Arabidopsis* centromeric regions, although largely heterochromatic, overall contain at least 47 expressed genes
- *Arabidopsis* contains several classes of proteins that are used in animal systems for processes not present in the plant, underscoring the idea that evolution makes use of the tools it is given to accomplish different tasks in different organisms
- Plants have evolved a host of signal transduction apparatus, perhaps to enable them to deal with their sessile nature
- *Arabidopsis* genome contains genes encoding RNA polymerase subunits not seen in other eukaryotic organisms



- *Arabidopsis* has genes unique to plants – approximately 150 unique protein families were found, including 16 unique families of transcription factors
- *Arabidopsis* has many gene families common to plants and animals which have been greatly expanded in plants – for instance, *Arabidopsis* contains 10-fold as many aquaporin (water channel) proteins than any other sequenced organism

